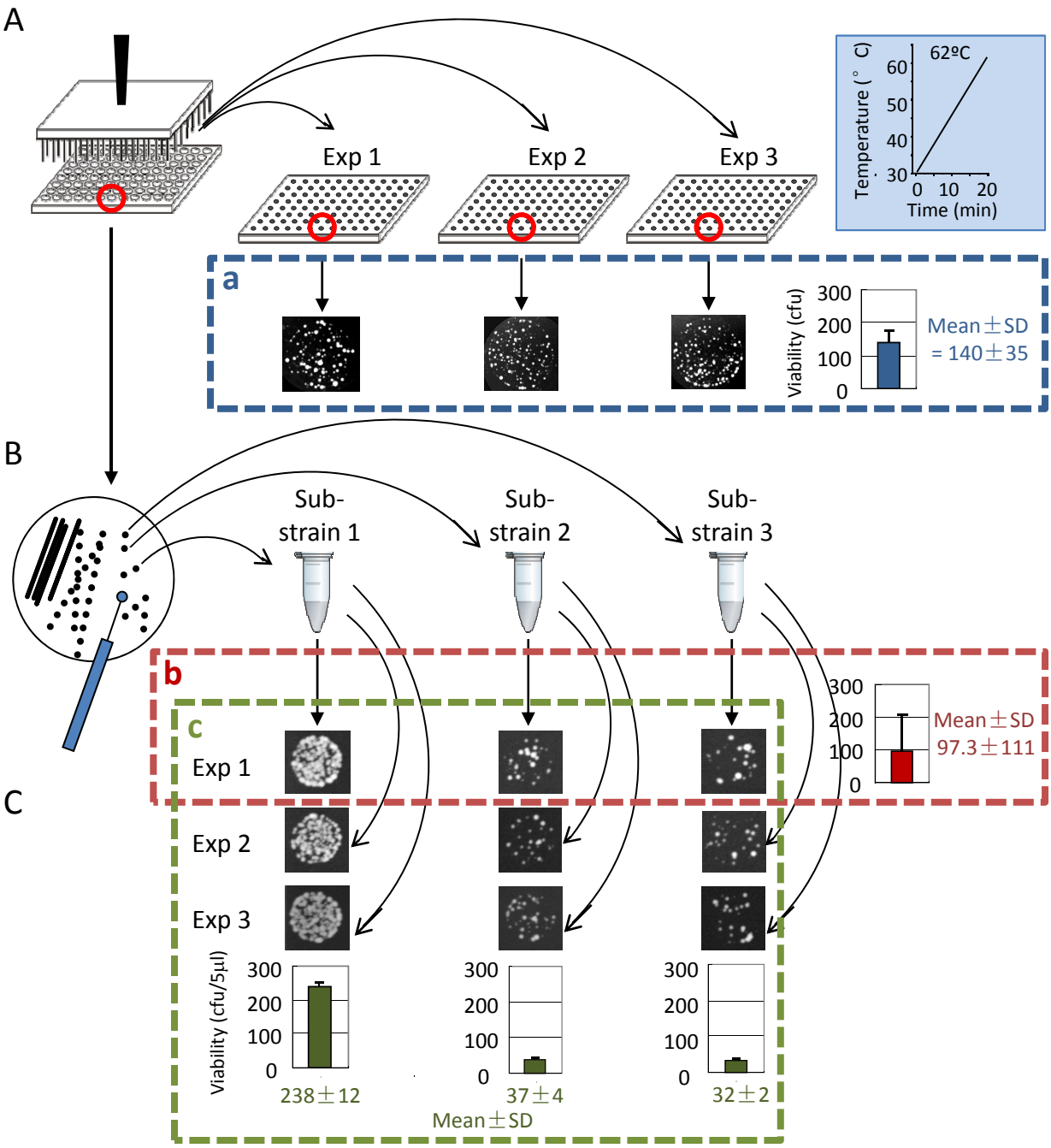


Figure S1. Screening strategy to detect stable variants within original yeast knockout (YKO) strains. Comparison of standard deviations (SD) from three independent experiments for each strain/substrain following heat-ramp stress (heat conditions as graphed). Related to Figure 1.

(A) The fourth verified YKO strain listed in the BY *MATa* collection (YAL061W) was pinned from the frozen archive without thawing, amplified on plates, and density-matched cultures were analyzed in the heat-ramp stress test in three independent experiments (box a).

(B) The same original YKO strain was streaked onto solid medium and three morphologically indistinguishable colonies were picked and tested as in panel A; variation between substrains derived from non-isogenic original YKO collection strain results in a large standard deviation (box b).

(C) In contrast, standard deviations are small for three independent experiments (Exp 1-3) on the same colony-derived substrain; for experiments 2 and 3, yeast were pinned from frozen stocks of the same original colonies in Exp 1, indicating that variation is biological rather than technical (box c).

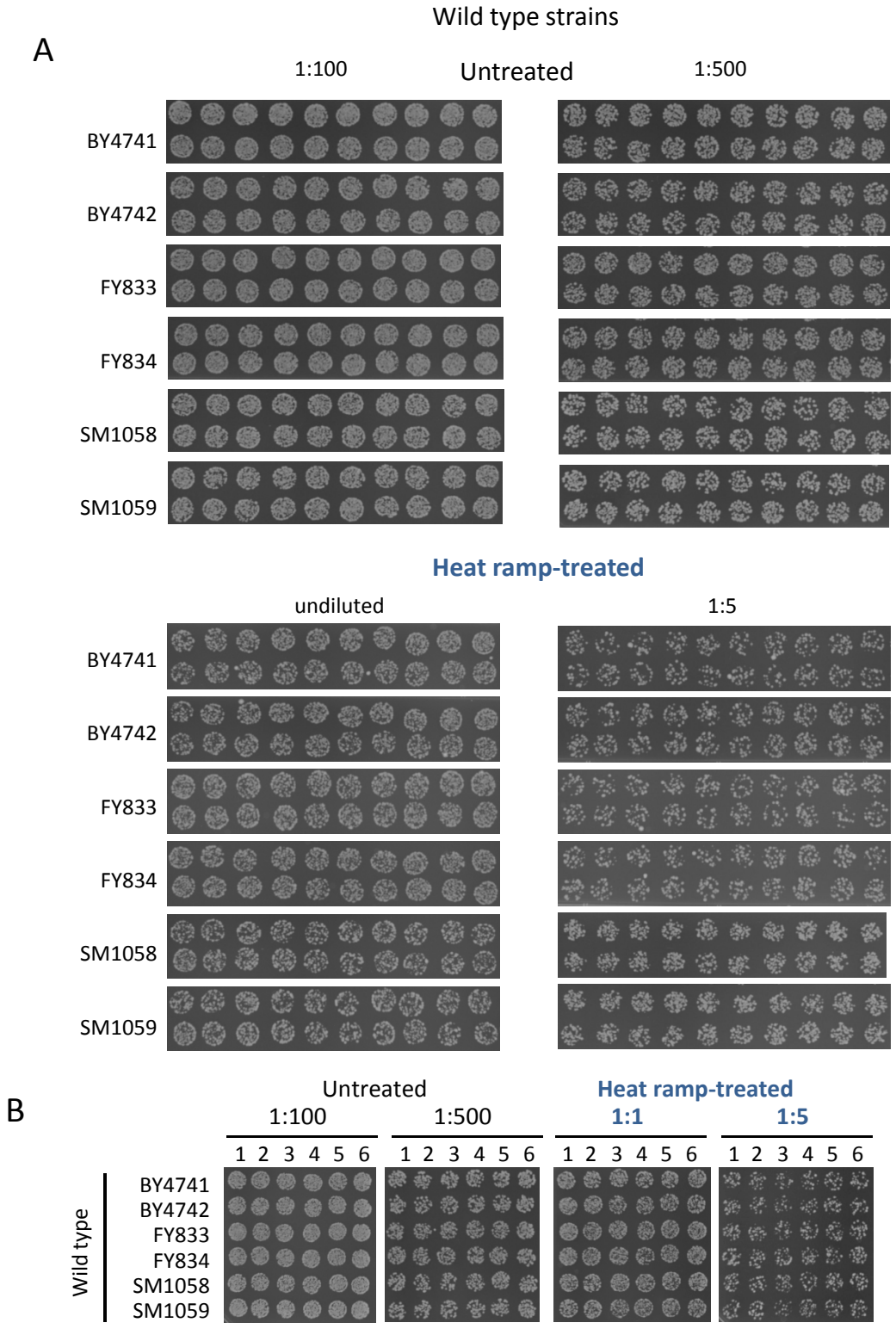
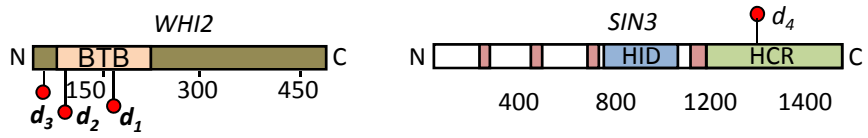


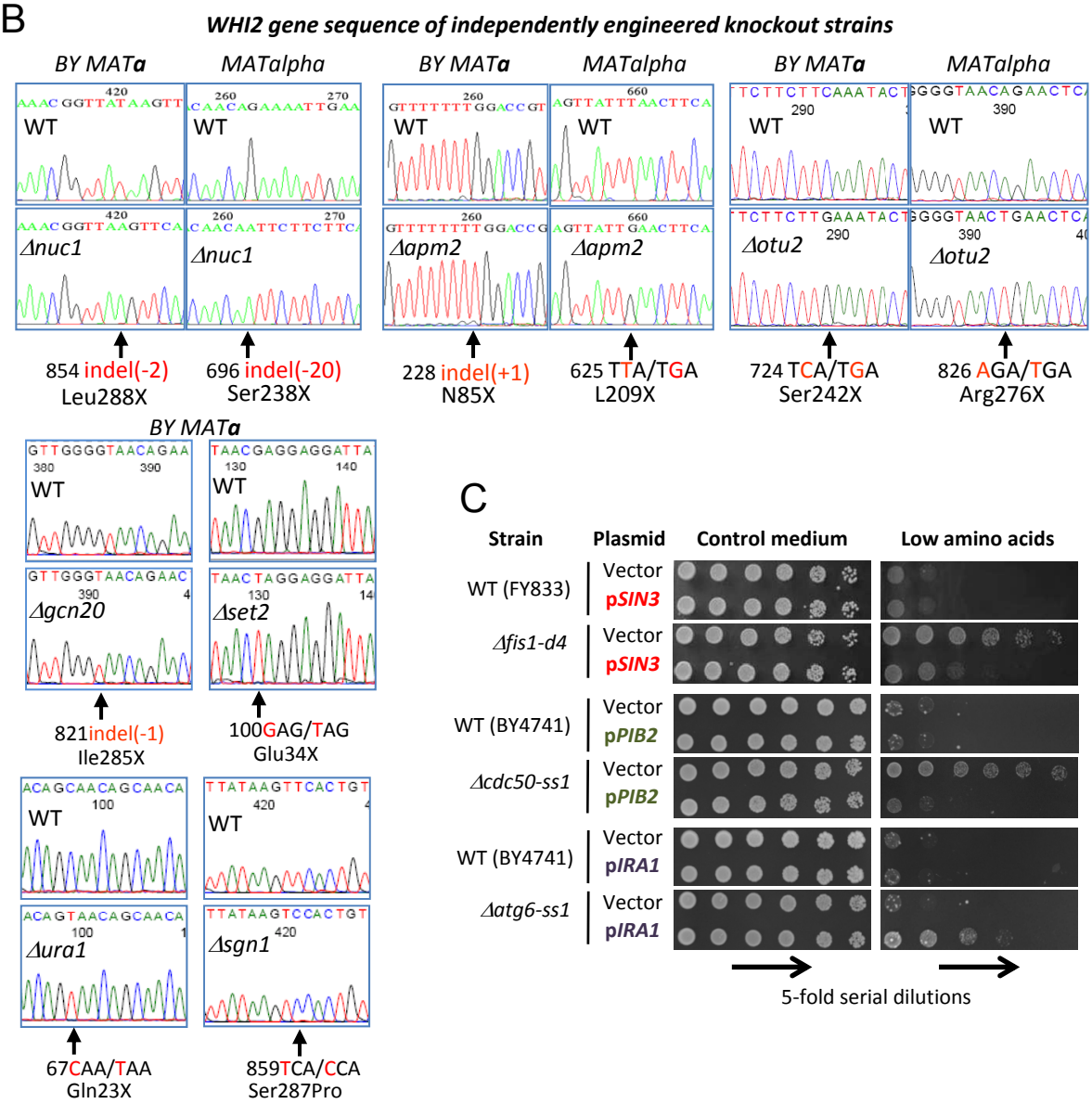
Figure S2. Invariant phenotypes of 26 substrains per wild type strain in the heat-ramp stress test.
(A) No variation was detected for colony-derived substrains #7-20 from each of 6 wild type strains in the heat-ramp stress test. Substrains #1-6 are shown in **Figure 2**.
(B) Dilutions of samples shown in **Figure 2C** confirm approximately equal starting cell numbers of wild type substrains #1-6 prior to treatment.

A



	Δfis1	WHI2	Protein	SIN3	Protein
C' gp 1	Δfis1d1	457GAG/TAG	Glu153X	WT	WT
	Δfis1d2	205TTA/TGA	Leu69X	WT	WT
	Δfis1d3	82TCA/TAA	Ser28X	WT	WT
C' gp 2	Δfis1d4	WT	WT	3694TAC/TAA	Tyr1232X

B



C

Figure S3. Identification of secondary mutations that cause overgrowth and heat-stress phenotypes.
(A) Summary of acquired secondary mutations found in four independently derived *FIS1* deletion strains.
(B) DNA sequence chromatograms of the *WHI2* gene in BY *MATa* and BY *MATalpha* YKO strains.
(C) Plasmids (p) expressing wild type genes rescue the corresponding secondary mutant genes from **Figures 5 and 6** in overgrowth assays, supporting a causal role for secondary mutations. Unexpectedly, the secondary mutation in *mgm1Δ* is dominant in both BY *MATa* and BY *MATalpha* YKOs independently of petite phenotype associated with deletion of *mgm1*, and therefore not testable here.

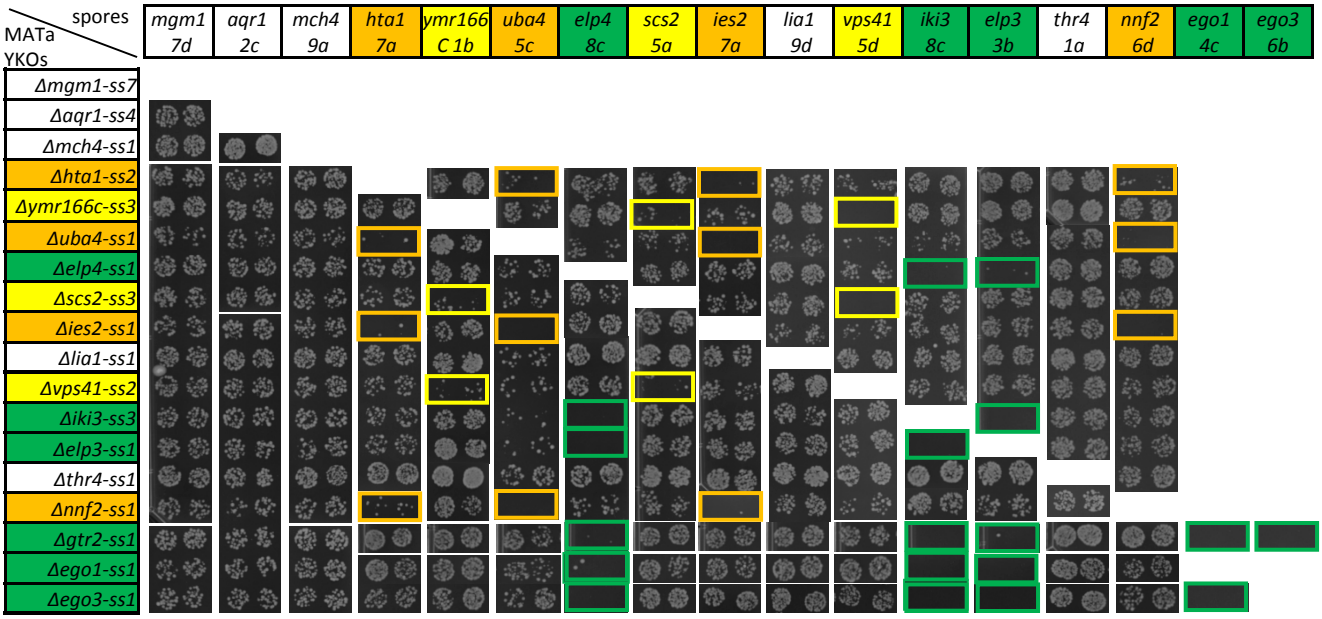
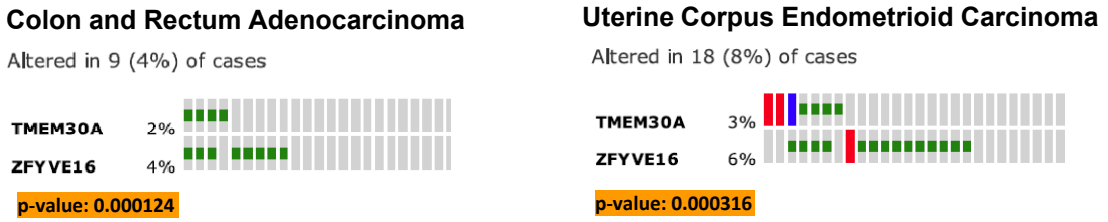
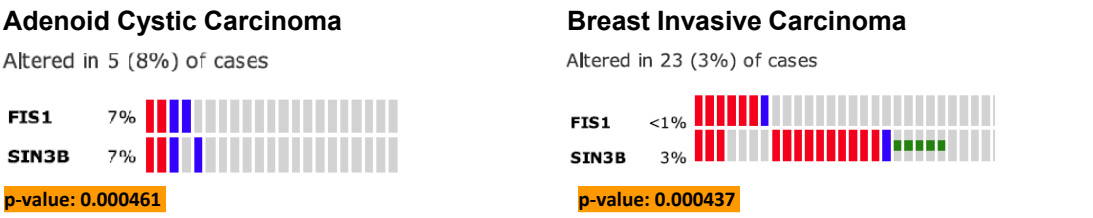


Figure S4. Complementation tests indicate that secondary mutations in 8 different molecular complexes (complementation groups) are responsible for the heat ramp-sensitive phenotypes in 18 different YKOs (BY *MATa*). Complementation test results are shown for those YKOs harboring secondary mutations responsible for a heat ramp-sensitive phenotype; these secondary mutations do not cause low amino acid overgrowth (strains listed in rows 19-36 of **Table S6**). Colony-derived substrains (left column) were mated with *MATalpha* ascospore segregants *(MATalpha his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0;* see **Table S8** for genotypes). The indicated strains bear the secondary mutation present in their respective BY *MATa* substrains (backcrosses described in **Figure 4**). Two diploid colonies for each cross were tested in the heat-ramp assay (shown above) and each was compared to wild type and to their respective parental haploid in the same assay (not shown). Strains bearing secondary mutations that belong to the same complementation group are color-coded. See **Figure 5C**.

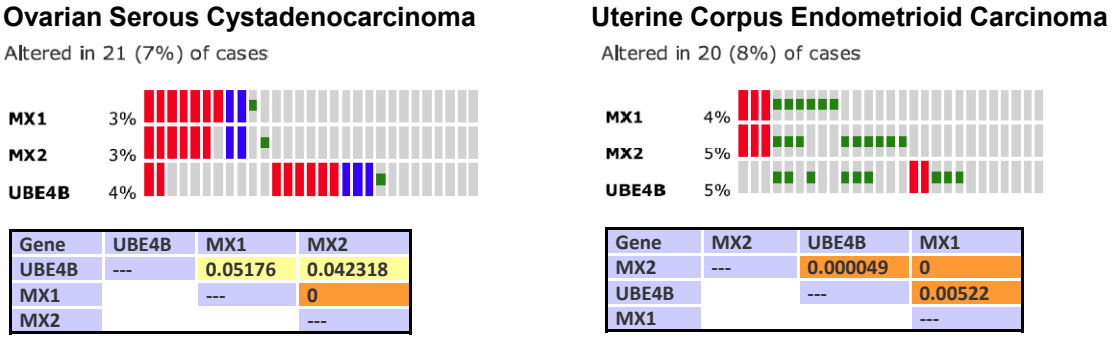
A



B



C



Gene	UBE4B	MX1	MX2
UBE4B	---	0.05176	0.042318
MX1		---	0
MX2			---

Gene	MX2	UBE4B	MX1
MX2	---	0.000049	0
UBE4B		---	0.00522
MX1			---

D

Uterine Corpus Endometrioid Carcinoma

Gene	KCTD19	KCTD9	SHKBP1	KCTD10	KCTD14	KCTD16	KCTD20	KCTD8	KCNRG	KCTD5	KCTD6
SETD2	0.000007	0.009659	0.000076	0.000015	0.000015	0	0.016185	0.000001	0.002205	0.03939	0.020797

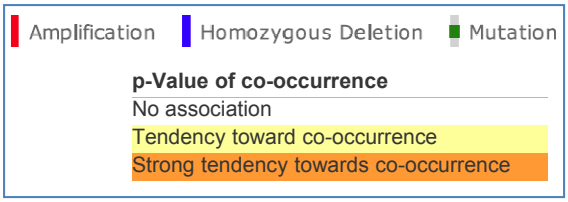


Figure S5. Individual tumors with co-occurrences of mutations in the human orthologs of both the yeast knockout and the co-occurring yeast secondary mutant genes.
(A) Mutations of human FIS1 and SIN3B co-occur in adenoid cystic and breast cancers.
(B) Mutations of TMEM30A and ZFYVE16 co-occur in colon and uterine cancers.
(C) Mutations of MX1/2 and UBE4B co-occur in ovarian and uterine cancers.
(D) p-Values for the co-occurrence of mutations in SETD2 with different KCTD family members in uterine cancers. Results shown are from the cBio Cancer Genomics Portal http://www.cbioportal.org/public-portal/index.do See **Figure 6E**.

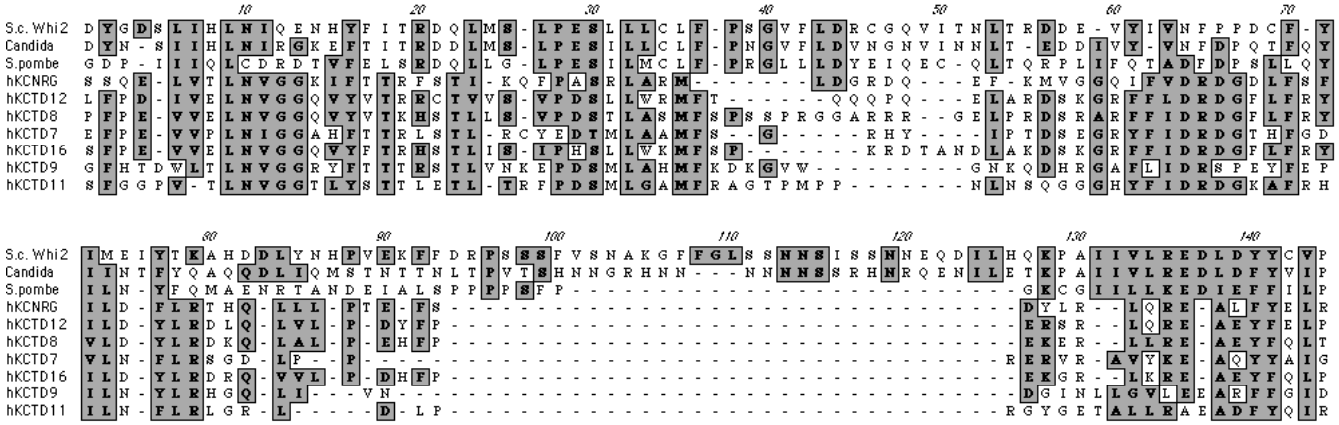


Figure S6. Yeast Whi2 shares significant amino acid sequence similarity to mammalian KCTD family proteins. Amino acid sequence alignment of the BTB/POZ domain of Whi2 from three species of yeast (top) compared to members of the human KCTD protein family (7 examples shown). See **Figure 6E**. The subclass of BTB/POZ domains found in Whi2/KCTD proteins is more closely related to the BTB-like tetramerization (T1) domain of voltage-gated potassium channels than to the BTB domains found in other proteins.

Supplemental Experimental Procedures

1. Strain list

Strain name	Genotype	Source
BY4741 [a]	*MATa his3 leu2 ura3 met15*	Brachmann et al., 1998
BY4742 [a]	*MATalpha his3 leu2 ura3 lys2*	Brachmann et al., 1998
SM1058	*MATa trp1 leu2 ura3 his4 can1*	Michaelis and Herskowitz, 1988
SM1059	*MATalpha trp1 leu2 ura3 his4 can1*	Michaelis and Herskowitz, 1988
FY833 [a]	*MATa his3 leu2 ura3 trp1 lys2*	Cerveny and Jensen, 2003
FY834 [a]	*MATalpha his3 leu2 ura3 trp1 lys2*	Cerveny and Jensen, 2003
BY4730	*MATa leu2 ura3 met15*	Brachmann et al., 1998
BY4739	*MATalpha leu2 ura3 lys2*	Brachmann et al., 1998
Δfis1-d1 [a]	*MATa his3 leu2 ura3 met15 fis1::KanMX4 whi2-1*	Cheng, et al., 2008
Δfis1-d2 [a]	*MATalpha his3 leu2 ura3 lys2 fis1::KanMX4 whi2-2*	Cheng, et al., 2008
Δfis1-d3 [a]	*MATalpha his3 leu2 ura3 trp1 lys2 fis1::HIS3 whi2-3*	Cheng, et al., 2008
Δfis1-d4 [a]	*MATa his3 leu2 ura3 trp1 lys2 fis1::URA3 sin3-1*	Cerveny and Jensen, 2003
d2	*MATalpha his3 leu2 ura3 lys2 whi2-2*	Cheng, et al., 2008
d4	*MATalpha his3 leu2 ura3 lys2 sin3-1*	this study; sporulated from WT BY4742 x *Δfis1d4*
Δcdc50_6a [a]	*MATa his3 leu2 ura3 cdc50::KanMX4 pib2-1*	this study; sporulated from WT BY4742 x *Δcdc50-ss1*
Δcdc50_6c [a]	*MATalpha his3 leu2 ura3 lys2 met15 cdc50::KanMX4 pib2-1*	this study; sporulated from WT BY4742 x *Δcdc50-ss1*
Δatg6_2c [a]	*MATalpha his3 leu2 ura3 lys2 atg6::KanMX4 ira1-1*	this study; sporulated from WT BY4742 x *Δatg6-ss1*
Δatg6_4d [a]	*MATa his3 leu2 ura3 lys2 met15 atg6::KanMX4 ira1-1*	this study; sporulated from WT BY4742 x *Δatg6-ss1*
Δmgm1_7c [a]	*MATalpha his3 leu2 ura3 mgm1::KanMX4 2nd-mut (HR^sLAA^o)*	this study; sporulated from WT BY4742 x *Δmgm1-ss7*
Δmgm1_7d [a]	*MATalpha his3 leu2 ura3 lys2 2nd-mut(HR^s)*	this study; sporulated from WT BY4742 x *Δmgm1-ss7*
Δcdc50_8d [a]	*MATalpha his3 leu2 ura3 lys2 pib2-1*	this study; sporulated from WT BY4742 x *Δcdc50-ss1*
Δdrs2_8a	*MATalpha his3 leu2 ura3 lys2 (2nd-mut HR^sLAA^o)*	this study; sporulated from WT BY4742 x *Δdrs2-ss2*
Δatp12_7c	*MATalpha his3 leu2 ura3 lys2 2nd-mut(HR^sLAA^o)*	this study; sporulated from WT BY4742 x *Δatp12-ss4*
Δalg9_3c	*MATalpha his3 leu2 ura3 lys2 2nd-mut(HR^sLAA^o)*	this study; sporulated from WT BY4742 x *Δalg9-ss9*
Δost4_11c	*MATalpha his3 leu2 ura3 lys2 2nd-mut(LAA^o)*	this study; sporulated from WT BY4742 x *Δost4-ss1*
Δvma6_9c	*MATalpha his3 leu2 ura3 lys2 2nd-mut(LAA^o)*	this study; sporulated from WT BY4742 x *Δvma6-ss1*
Δthr4_1a	*MATalpha his3 leu2 ura3 lys2 2nd-mut(HR^sLAA^o)*	this study; sporulated from WT BY4742 x *Δthr4-ss1*
Δaqr1_2c	*MATalpha his3 leu2 ura3 lys2 aqt1::KanMX4 2nd-mut(HR^s)*	this study; sporulated from WT BY4742 x *Δaqr1-ss4*
Δmch4_9a	*MATalpha his3 leu2 ura3 lys2 2nd-mut(HR^s)*	this study; sporulated from WT BY4742 x *Δmch4-ss1*
Δhta1_7a	*MATalpha his3 leu2 ura3 lys2 hta1::KanMX4 2nd-mut(HR^s)*	this study; sporulated from WT BY4742 x *Δhta1-ss2*
Δnnf2_6d	*MATalpha his3 leu2 ura3 lys2 2nd-mut(HR^s)*	this study; sporulated from WT BY4742 x *Δnnf2-ss1*
Δymr166c_1b	*MATalpha his3 leu2 ura3 lys2 ymr166c::KanMX4 2nd-mut(HR^s)*	this study; sporulated from WT BY4742 x *Δymr166c-ss3*
Δuba4_5c	*MATalpha his3 leu2 ura3 lys2 2nd-mut(HR^s)*	this study; sporulated from WT BY4742 x *Δuba4-ss1*
Δelp4_8c	*MATalpha his3 leu2 ura3 lys2 2nd-mut(HR^s)*	this study; sporulated from WT BY4742 x *Δelp4-ss1*
Δscs2_5a	*MATalpha his3 leu2 ura3 lys2 2nd-mut(HR^s)*	this study; sporulated from WT BY4742 x *Δscs2-ss3*
Δies2_7a	*MATalpha his3 leu2 ura3 lys2 2nd-mut(HR^s)*	this study; sporulated from WT BY4742 x *Δies2-ss1*
Δlia1_9d	*MATalpha his3 leu2 ura3 lys2 lia1::KanMX4 2nd-mut(HR^s)*	this study; sporulated from WT BY4742 x *Δlia1-ss1*
Δvps41_5d	*MATalpha his3 leu2 ura3 lys2 vps41::KanMX4 2nd-mut(HR^s)*	this study; sporulated from WT BY4742 x *Δvps41-ss2*
Δiki3_8c	*MATalpha his3 leu2 ura3 lys2 2nd-mut(HR^s)*	this study; sporulated from WT BY4742 x *Δiki3-ss3*
Δelp3_3b	*MATalpha his3 leu2 ura3 lys2 2nd-mut(HR^s)*	this study; sporulated from WT BY4742 x *Δelp3-ss1*
Δego1_4c	*MATalpha his3 leu2 ura3 lys2 ego1::KanMX4 2nd-mut(HR^s)*	this study; sporulated from WT BY4742 x *Δego1-ss1*
Δego3_6b	*MATalpha his3 leu2 ura3 lys2 ego3::KanMX4 2nd-mut(HR^s)*	this study; sporulated from WT BY4742 x *Δego3-ss1*

[a] Whole genome sequenced strains

2nd-mut HR^s The strain harbors a 2nd mutation responsible for the heat-ramp sensitive phenotype

2nd-mut LAA^o The strain harbors a 2nd mutation responsible for the low amino acid overgrowth phenotype

2nd-mut HRs LAA^o The strain harbors a 2nd mutation responsible for both the heat-ramp sensitive and the low amino acid overgrowth phenotypes

2. Mathematic model for estimating heterogeneity of the YKO collection

Scoring heterogeneity

Let *K* substrains be picked from a given yeast knockout strain. Each of the *K* substrains is subjected to heat ramp stress. The phenotype ('-', '+', '++','+++' or '++++') of each substrain was transformed into a numeric score between 0 and 4 respectively. Thus the measured heterogeneity of the *i-th* parental knockout strain is quantified by a set of K_i scores. Let s_i be the most frequently occurring score for the *i-th* parental stain. Substrains are labeled as "prevalent" or "variant" as follows: If the score is equal to the most frequent score (among the K_i scores*)* the substrain is designated "prevalent" and is labeled with a '0', otherwise the substrain is designated "variant" and is labeled with a '1'. A parental knockout strain that is heterogeneous is expected to produce at least one variant substrain (for sufficiently large *K)* while parental strains that have no variant substrains are designated homogeneous parental strains.

Statistical model

For simplicity assume that there are only two types of parental knockout strains: 1) parental strains that can never have variants (i.e. respond homogeneously to heat ramp stress) and 2) parental strains that can have variant substrains. In the latter case, we make the further assumption that all heterogeneous parental strains are identical in the sense that they all have the same probability of observing a variant substrain. Of course this is not the case in reality, but it is a useful assumption for estimating the proportion of heterogeneous parental strains.
Let the heterogeneity of a strain be labeled with a stochastic variable *c,* where c=1 corresponds to heterogeneous and where c=0 corresponds to homogeneous. Let *q* be the probability that a randomly selected parental strain is heterogeneous, i.e.

$$p(c)=\begin{Bmatrix} 1-q & if\ c=0 \\ q & if\ c=1 \end{Bmatrix} \quad (1)$$

A measurement consists of *K* Bernoulli trials, where a 0 (failure) indicates a *prevalent* substrain while a 1 (success) indicates an *variant* substrain, in other words,

$$p(k|c=0,K)=\begin{Bmatrix} 1 & if\ k=0 \\ 0 & otherwise \end{Bmatrix} \quad (2)$$

Based on the reproducibility of independent experiments, we assume that the noise in the measurements is so low that homogenous parental strains never produce measurements with variant substrains. On the other hand, for heterogeneous parental strains (c=1), the number of variant strains in the measurement (the number of 1's) is binomially distributed, i.e.

$$p(k|c=1,K)=\frac{K!}{k!(K-k)!}p^k(1-p)^{K-k} \quad (3)$$

The simplest model of the knockout collection is to assume that all heterogeneous parental knockout strains have the same underlying variability *p.* Putting this all together, if we select a parental strain at random and perform a measurement over *K* substrains, then probability of observing exactly *k* variant substrains is:

$$P(k|K,p,q)=((1-q)+q(1-p)^K)\delta_{k,0}+(1-\delta_{k,0})q\frac{K!}{k!(K-k)!}p^k(1-p)^{K-k} \quad (4)$$

Where the indicator function $\delta_{k,0}$ is 1 if k=0, and 0 otherwise.

Detection of heterogeneous parental strains

The sensitivity of a measurement (the probability of detecting variation in a parental strain that is actually heterogeneous) depends on the number of substrains that are sampled (*K)* as well as the proportion of variant substrains (*p*). We say that a heterogeneous strain is "detected" if at least one substrain is a variant, i.e. *k*≥1. The probability of detecting at least one variant substrain given a heterogeneous parental strain (i.e. c=1) can be derived from equation (3). The detection probability as a function of *K* is shown in **Fig. SM1**.

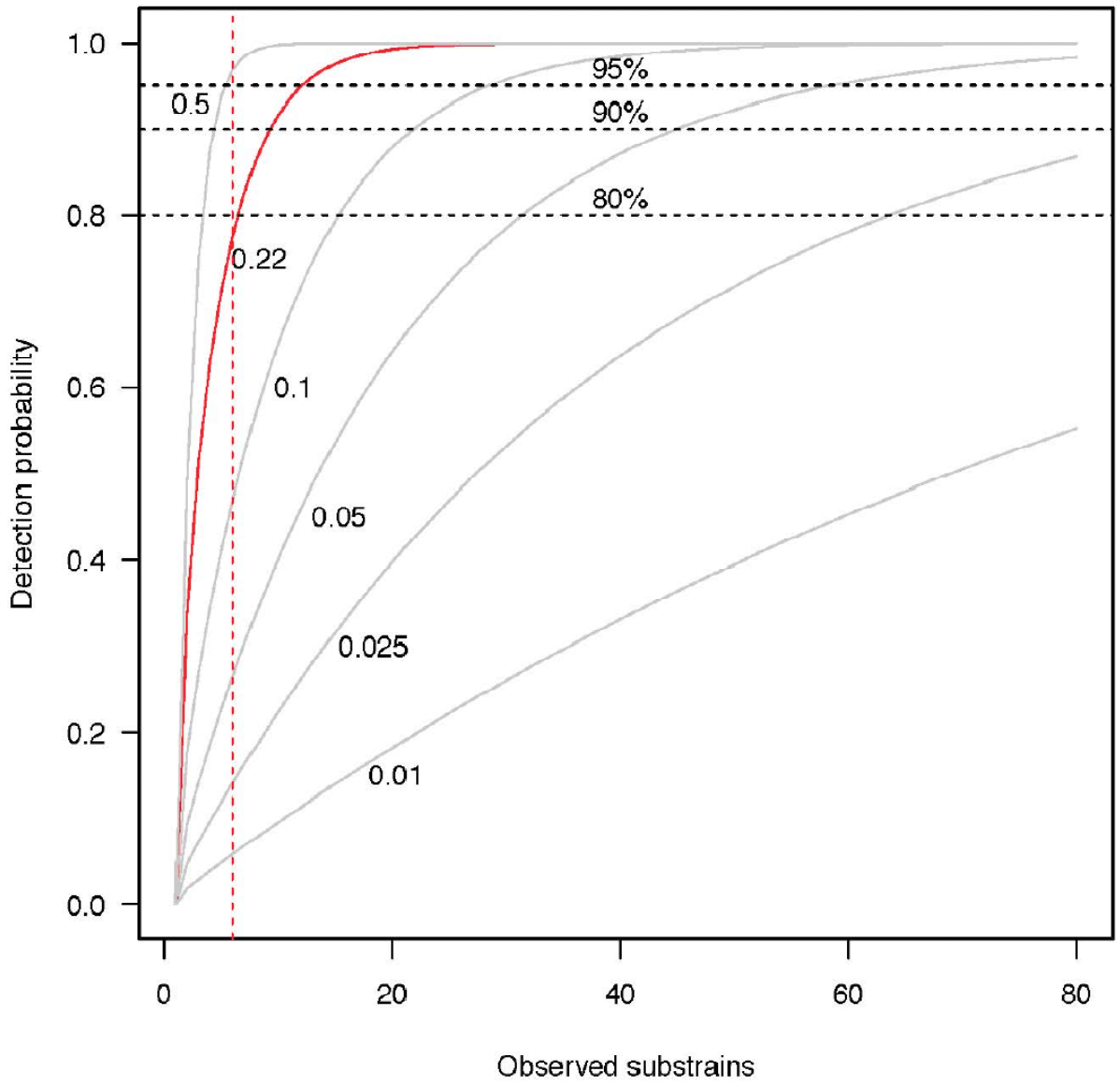
Estimation of model parameters

We estimate the proportion of heterogeneous strains (*q*) as well as the proportion of heterogeneous substrains (*p)* in a heterogeneous strain by maximizing the likelihood of the data given the *p* and *q.* The data consists of a set of *N* experiments.

For each experiment, we record the number of substrains that are tested (K_i) and the number (k_i) of substrains that have scores that deviate from the most frequent score. To estimate *p* and *q* we maximize the log likelihood

$$\log(L(p,q|data))=\sum_{i=1}^{N}\log(P(k_i|K_i,p,q)) \quad (5)$$

Using the data from 250 randomly selected samples yields q= 56% and p=22%. The maximum likelihood estimate is performed by the R script **ml.R** using data in the file **data.txt**.



Graph. Detection probability as a function of observed substrains. Different curves are labeled by the amount of true variability (*p*) of a parental strain. The vertical dotted line corresponds to a measurement with 6 substrains. The red curve corresponds to *p=0.22* (estimated from our experiment with 250 randomly selected parental strains). In this case we see that the detection probability is roughly 78%. To achieve 95% detection probability of

detecting that a parental strain is heterogeneous or not, would require measurements of roughly 15 substrains. The figure is generated by the R script **"detection.R".**

3. Calculations for significance of finding similar phenotypes caused by secondary mutations in independently constructed YKOs.

Heat-ramp sensitivity assay.

Based on data for 250 randomly selected YKOs from **Figure 2**, there are 116 YKOs with heat-sensitive phenotypes (having at least one heat-sensitive substrain, with or without substrain variation) while the remaining 134 YKOs do not have this type of phenotype (**Table S1**). Among the 116 YKOs, 73.7 (116×63.5%, 63.5% is extracted from **Figure 4C**) have this heat-sensitive phenotype due to a secondary mutation. For a randomly selected knockout strain with an engineered gene KO that does not cause a heat-sensitive phenotype, there will be two possibilities. (1) There is no secondary mutation (no heat-sensitive phenotype), or (2) there is a secondary mutation causing a heat-sensitive phenotype; the occurrence of these two possibilities is 134:73.7, respectively. On this basis, when 36 YKOs (from **Figure 5B**) in which the gene KO itself does not cause the heat-sensitive phenotype are tested, the expected proportion for having no secondary mutation or having a secondary mutation causing heat-sensitive phenotype shall be 23.2:12.8. However, the observed proportion is 13:23 (23 from **Figure 5B**).

	No HR^S phenotype	HR^S due to 2^{nd} mutations
Observed/ inferred from 250 random YKOs	134	73.7
Expected if 36 random YKOs are tested	23.2	12.8
Observed when 36 YKOs tested	13	23

An exact binomial test (2-sided) rejects the null hypothesis that the observed proportion (13:23) from the 36 YKOs tested happens by chance with a significance *p*-value = 0.0006925. So it is highly significant that the independently constructed YKOs deleted of the same genes tend to evolve similar secondary mutations that are responsible for the heat-sensitive phenotype.

Low amino acid overgrowth phenotype

Based on the whole library screen, there are 749 YKOs (among a total 4,847 BY *MATa* YKOs) that have an overgrowth phenotype on low amino acid (LAA) medium (**Table S4**) while the

remaining 4,096 YKOs do not have this phenotype. Among these 749 YKOs, 476.9 (749×63.5%) may have this phenotype due to the secondary mutations. Similarly as above:

	No LAA^O phenotype	LAA^O due to 2^{nd} mutation
Observed/ inferred from total 4847 YKOs	4096	476.9
Expected if 23 random YKOs are tested	20.6	2.4
Observed when 23 YKOs tested	7	16

An exact binomial test (2-sided) rejects the null hypothesis that the observed proportion (7:16, from **Figure 5B**) from the 23 YKOs tested happens by chance with a significant *p*-value = 2.351e-11. So it is highly significant that the independently constructed YKOs deleted of the same genes tend to evolve similar secondary mutations that are responsible for the low amino acid overgrowth phenotype.

Combined heat-ramp sensitive and low amino acid overgrowth phenotypes

Based on the whole library screen, there are 399 YKOs (among total 4,847 BY *MATa* YKOs) having at least one substrain showing both the heat-ramp sensitive and low amino acid overgrowth phenotype, while the remaining 4,448 YKOs do not have such combined phenotypes. Among these 399 YKOs, 253.4 (399×63.5%) may have this phenotype due to secondary mutations. Similarly as above:

	No HR^S/LAA^O	HR^S/LAA^O due to 2^{nd} mutation
Observed/ inferred from total 4847 YKOs	4448	253.4
Expected if 19 random YKOs are tested	18.0	1.0
Observed when 19 YKOs tested	6	13

An exact binomial test (2-sided) rejects the null hypothesis that the observed proportion (6:13) from the 19 YKOs tested happens by chance with a significant *p*-value = 4.778e-13. So it is highly significant that the independently constructed YKOs deleted of the same genes tend to evolve similar secondary mutations that are responsible for the combined heat-ramp sensitive and amino acid growth phenotype.

4. Selection of strains used in this study.

The 250 original YKO strains in **Figure 2** were selected using a random number generator.

The 749 original YKO strains in **Figure 3** were identified in a genome-wide screen for low amino acid overgrowth, testing pinned original strains (see **Experimental Procedures**).

The rationale for testing different numbers of substrains derived from the 749 in **Figure 3C** are as follows: YKOs the strongest (3-tested) and moderate overgrowth phenotypes (6-tested) in the genome-wide screen, and the 120 YKOs with 9-tested include 111 YKOs with 6 additional colonies analyzed after it was found that the first 3 tested substrains lacked overgrowth phenotypes, and 9 additional YKOs from the 749 selected for a separate study. These groups do not overlap and were grouped without regard to substrain variation.

The 69 YKO strains in **Figures 4 and 5** were selected arbitrarily by compiling strains with tetrads produced from all other past/present projects in the Hardwick laboratory. We determined subsequently that among these 69 original *MATa* YKOs, 49% have variant substrains (similar to the randomly selected strains, see **Figure 2**), and that these strains are distributed across the genome. From confidential records, it was determined that the relevant 40 YKOs with secondary mutations (**Figures 4 and 5**) were constructed by 15 of the 16 participating laboratories, and that the most relevant subset of 26 YKO strains were constructed by 11 different laboratories from 6 different countries, as if they had been randomly selected from the entire collection. Although >1,000 of the ~5,000 YKOs in the BY *MATa* collection have heat ramp-resistant phenotypes (Teng et al. 2011), convenient for this study most all of these strains have secondary mutations that decrease (not increase) cell survival following stress, helping to dispel any concerns that these strains have selected general survival mutations.

5. Direct whole genome sequencing

Starting with 1-2 mg of gDNA, samples were sheared by sonication on the Covaris S2 System to a fragment size of approximately 165 bp with the following conditions: Duty Cycle 10%, Intensity 5%, Cycles/Burst 200, Time 60 seconds and Number of cycles 6. Libraries were constructed according to the protocol provided in the NEBNext DNA Sample Prep Master Mix Set 3 using the SOLiD Fragment Library Barcoding Kit Module 1-16 utilizing barcodes 1-8. Assessment of the yield and size distribution of the amplified library was performed on the Agilent 2100 Bioanalyzer using the High Sensitivity Chip. Size selection and excess adaptor and primer/dimer removal of the samples were carried out by running them on a 1% TAE agarose gel. Libraries were excised between 200-300 bp size range and gel purified using the Invitrogen Purelink Gel Extraction Kit. Samples were once again run on the Agilent 2100 Bioanalyzer with the High Sensitivity Chip to assess the yield and size distribution of the gel purified library. Quantification of the libraries was performed by qPCR using the TaqMan Gene Expression Assay as outlined in the Applied Biosystems SOLiD Library Preparation Guide.

Libraries were pooled to a final concentration of 500 pM and emulsion PCR performed to generate 708 million beads to deposit on a Full Slide. Sequencing was performed on the SOLiD 4 System using a single read, 50 bp fragment run.

References:

Michaelis, S. and Herskowitz, I. (1988). The a-factor pheromone of Saccharomyces cerevisiae is essential for mating. Mol. Cell. Biol. 8:1309-1318.

Cerveny, K.L. and Jensen, R.E. (2003). The WD-repeats of Net2p interact with Dnm1p and Fis1p to regulate division of mitochondria. Mol Biol Cell. 14: 4126-4139.